Quantification of tumor infiltrating lymphocytes (TILs) from pathology slides reflects molecular immune phenotypes

Ciyue Shen[1], Shima Nofallah[1], Jake Conway[1], Chintan Parmar[1], Michael Drage[#], Fedaa Najdawi[1], Darpan Sanghavi[1], Limin Yu[1], Raymond Biju[1], Daniel Borders[1], Matthew Bronnimann[1], Laura Chambre[#], Isaac Finberg[1], Jonathan Glickman[#], Michael Griffin[1], Sidharth Gupta[#], Natalia Harguindeguy[1], Nhat Le[1], Stephanie Hennek[1], Syed Ashar Javed[1], Christian Kirkup[1], Miles Markey[1], Michael Nercessian[1], Daniel Shenker[1], Sandhya Srinivasan[#], Vignesh Valaboju[1], Samuel Alberto Vilchez Mercedes[1], Bahar Rahsepar[1], Ryan Leung[1], Archit Khosla[#], Benjamin Glass[1], Amaro Taylor-Weiner[#], Ylaine Gerardin[1], John Abel[1]

[1]PathAI, Boston, MA
[#]Employed by PathAI at time of study

PathAI
Abstract #122

BACKGROUND

Examination of histopathology slides is a crucial step in making cancer diagnoses and treatment decisions. Rapid developments in machine learning (ML) models in digital pathology have enabled quantitative high-resolution information to be extracted from whole-slide images (WSIs). Meanwhile, genomic tests and molecular assays have also become powerful in assisting pathologists and oncologists in decision making. However, these tests are not routinely performed, leading to inconsistent availability of molecular information. Prior studies have shown that actionable molecular phenotypes can be predicted directly from hematoxylin and eosin (H&E)–stained slides.[1] Here, we show that tissue and cell classification models deployed on WSI of H&E-stained tissue, extracted human-interpretable features (HIFs) quantifying the tumor microenvironment, and investigated the association between abundance and distribution of tumor infiltrating lymphocytes (TILs) and molecular phenotypes.

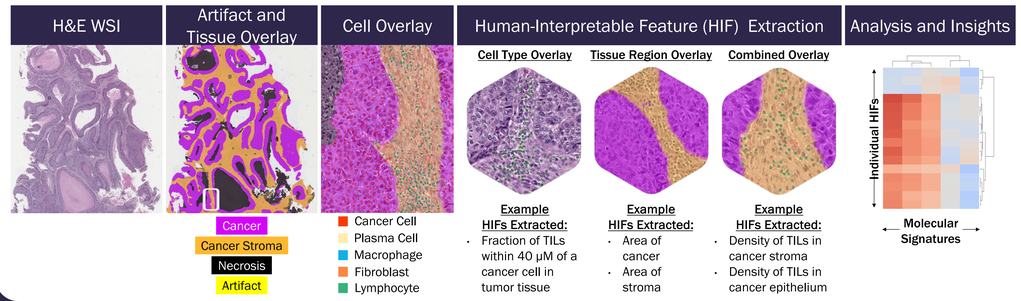


Figure 1. PathExplore model overview. PathExplore cell and tissue models are deployed on H&E-stained WSI to predict regions of cancer, stroma, and necrosis, as well as cell types including cancer cells, fibroblasts, lymphocytes, macrophages, and plasma cells. From these predictions, HIFs can be extracted and used for downstream analyses examining their relation to clinically relevant parameters (e.g., gene signatures).

CONCLUSIONS

- Histopathology image-based quantification of TILs is consistently associated with immune phenotypes derived from molecular measurements.
- Results shown here suggest that quantitative HIFs extracted from tissue and cell classification models provide rich information for understanding of inflammation in the tumor microenvironment and potential discovery of immune biomarkers.

METHODS

Machine Learning Models. Cell and tissue classification models based on convolutional neural networks were trained to identify and quantify features in the tumor microenvironment (TME) using H&E slides with annotations collected from US board-certified pathologists resulting in PathExplore models specific for eight indications: breast cancer, colorectal cancer, gastric cancer, melanoma, non-small cell lung cancer, pancreatic cancer, prostate cancer, and renal cell carcinoma. Model performance was evaluated by comparison to expert pathologist annotations using patch sampling as described in Gerardin et al.[2]

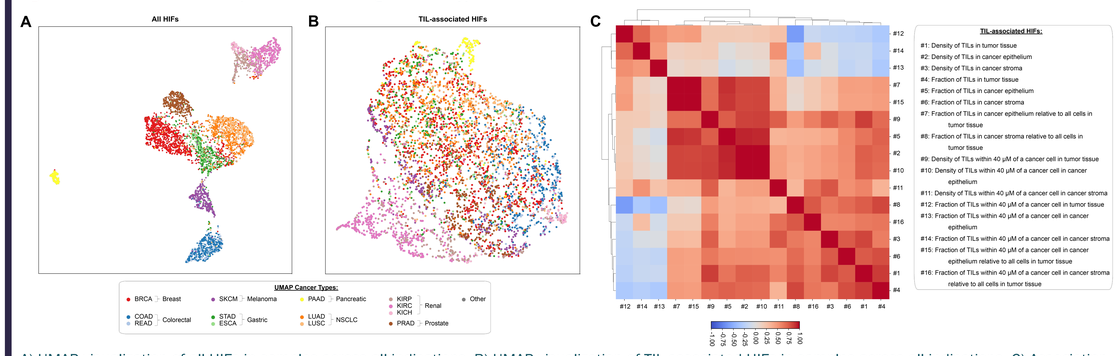
Model Deployment. PathExplore models were deployed on >5,000 WSI from 13 cancer types obtained from the Cancer Genome Atlas (TCGA). HIFs were extracted from cell and tissue predictions describing histological features including cell types, tissue regions, and the relationships between these features (Figure 1). Samples with missing values, utilized for purposes other than primary diagnosis, or determined to be non-cancerous were excluded; samples with outlier HIFs were manually inspected for model and slide quality and excluded as needed. A total of >4,700 samples were included in the final analysis (BRCA, N=1028; COAD, N=438; READ, N=151; STAD, N=379; ESCA, N=176; SKCM, N=329; LUAD, N=469; LUSC, N=424; PAAD, N=144; PRAD, N=404; KIRP, N=248; KIRC, N=471; KICH, N=106). Raw features (e.g., total tumor area, total cancer cell count) were excluded from downstream analyses, resulting a total of ~300 HIFs describing proportional areas, proportional counts, and density ratios. Sixteen HIFs were selected to represent TME properties relating to TIL abundance and distribution ("*TIL-associated HIFs*"; Figure 2).

Exploratory analyses. Uniform manifold approximation and projection (UMAP) was performed to compare HIFs between cancer types. HIFs were z-scored across all cancer types for standardization. UMAP was parameterized with 15 neighbors, an embedding dimension of 2, and the Euclidean distance metric. The intercorrelation among the TIL-associated HIFs were evaluated using Spearman correlation. Associations between TIL-associated HIFs and gene expression[3] were evaluated using Spearman correlation. Associations between TIL-associated HIFs and publicly available immune signatures[4,5] and data were evaluated using Spearman correlation. LASSO logistic regression models were used to predict median-binarized gene expression as well as immune phenotypes derived from immune signatures. For all regression models, 5-fold cross-validation was performed on individual indications and pan-cancer and AUROC values were reported.

RESULTS

- Model predicted HIFs were compared between indications and to each other. Assessing all HIFs, high levels of separation were observed between samples of distinct cancer types (Figure 2A). Within a cancer type, some separation was observed among subtypes (e.g., between KIRP, KIRC, KICH). Features associated with TILs did not yield the same degree of separation between samples of different cancer types (Figure 2B). Certain TIL-associated HIFs were observed to be correlated or anti-correlated with each other (Figure 2C).
- TIL-associated HIFs, such as the fraction of TILs in cancer epithelium (cTIL fraction), were correlated with gene expression[3] of known lymphocyte markers (Figure 3A), including CD8A (median Spearman ρ = 0.54 for individual indications). Regularized regression models using a panel of TIL-associated HIFs accurately predicted median-binarized expression of CD3G (Figure 3B), CD2 (Figure 3C) and CD8A (Figure 3D) (median AUROC 0.748-0.756 for individual indications, pan-cancer AUROC 0.777-0.781), with best performance in melanoma (AUROC 0.788-0.824) and breast cancer (AUROC 0.799-0.819).
- cTIL fraction was found to be correlated with immune signature scores derived from gene expression, including a published lymphocyte infiltration signature score[4] (Figure 4A; ρ = 0.497) and T-cell signature score[5] (Figure 4B; ρ = 0.470). In addition, classification models using TIL-associated HIFs can predict the inflammatory subtype (Figure 4C; C3 subtype[4], median AUROC = 0.685, pan-cancer AUROC = 0.766 ± 0.019, 5-fold cross-validation) and the immune-enriched non-fibrotic subtype (Figure 4D; IE subtype[5], median AUROC = 0.749, pan-cancer AUROC = 0.737 ± 0.014).

Figure 2. Examination of PathExplore HIFs across eight cancer types.

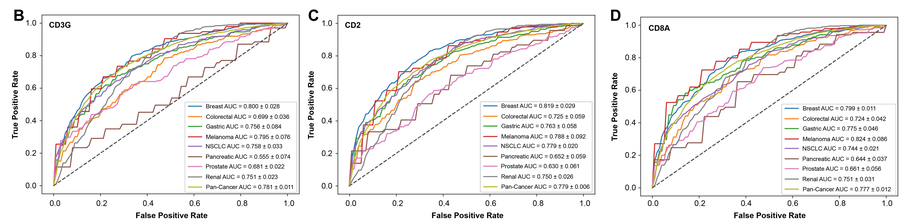


A) UMAP visualization of all HIFs in samples across all indications. B) UMAP visualization of TIL-associated HIFs in samples across all indications. C) Association between TIL-related HIFs was assessed via unsupervised clustering of pairwise Spearman correlations.

Figure 3. Association between TIL-associated HIFs and gene expression[3].

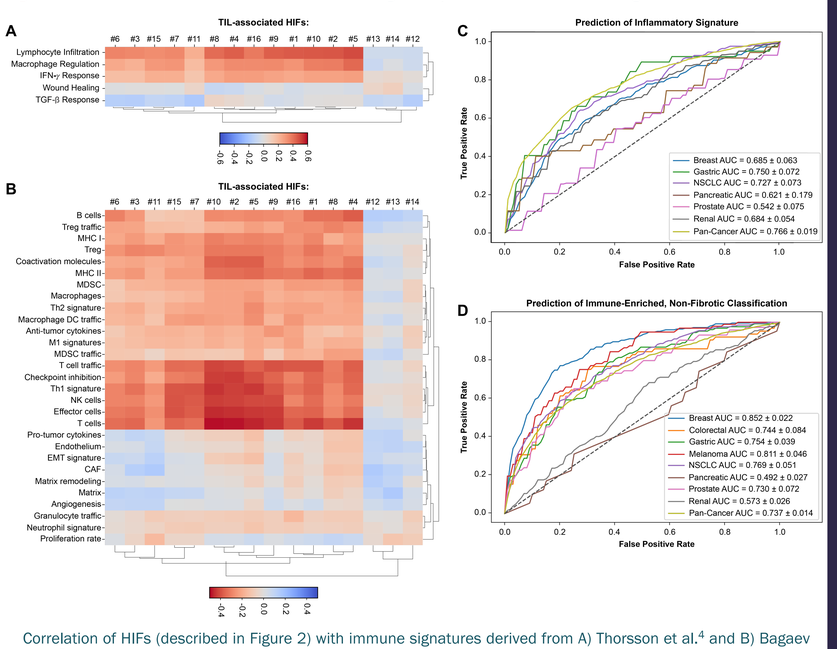
A

Gene	Median Spearman ρ (range)
CD3D	0.542 (0.218, 0.654)
CD2	0.540 (0.247, 0.665)
CD8A	0.539 (0.225, 0.657)
NKG7	0.539 (0.150, 0.648)
CD3G	0.537 (0.244, 0.656)
CD3E	0.533 (0.242, 0.642)
TBX21	0.528 (0.212, 0.625)
CCL5	0.523 (0.208, 0.654)
UBASH3A	0.523 (0.194, 0.651)
TRBC2	0.521 (0.100, 0.643)



A) The top 10 genes after sorting by median Spearman correlation with cTIL fraction across all eight cancer types. ROC curves depicting performance of regression models, developed using a panel of TIL-associated HIFs, to predict B) CD3G, C) CD2, and D) CD8A expression.

Figure 4. Association between TIL-associated HIFs and gene expression-derived immune signature scores.



Correlation of HIFs (described in Figure 2) with immune signatures derived from A) Thorsson et al.[4] and B) Bagaev et al.[5] C-D) ROC curves depicting the performance of classification models trained using TIL-associated HIFs to predict C) inflammatory subtype or D) immune-enriched, non-fibrotic subtype.

ACKNOWLEDGMENTS: The results shown here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. This poster template was developed by SciStories LLC. https://scistories.com

REFERENCES:
1. Diao JA, et al. Nat Comm. 2021; 12(1):1613
2. Gerardin Y et al., 2023 arXiv:2306.04709
3. Goldman MJ. Et al. Nat Biotechnol. 2020. 38, 675-678
4. Thorsson A. et al. Immunity. 2018; 48-4: 812-830
5. Bagaev A. et al. Cancer Cell. 2021; 39: 845-865

CONTACT: john.abel@pathai.com
publications@pathai.com

Society for Immunotherapy of Cancer 38th Annual Meeting, November 1-5, 2023 San Diego, CA